

**СУ “Св. Климент Охридски”
Стопански факултет
Управленски информационни системи**

**Курсова работа
на тема**

Управление на ресурсите на данни

**Катерина Върбанова
Фак. № 10711**

**София
2012**

Съдържание:

I Въведение:	2
II Теоретична част:	2
1. Терминология	2
2. Проблеми на традиционната файлова организация на данните	2
3. Управлението на данни с помощта на БД подхода	3
4. Системи за управление на Бази Данни.....	3
5. Компоненти на една Business Intelligence среда.....	4
5.1 Бизнес моделиране.....	4
5.2 Моделиране на данните.....	5
5.3 Online Transaction Processing.....	6
5.4 ETL, Data Warehouse, Data mart.....	7
5.5 Online Analytical Processing.....	7
III Практическа част:	8
Казус.....	8
Business view.....	10
Process view.....	11
Applications view.....	12
Technical view.....	12
Алтернативен вариант.....	13

I Въведение:

Все повече компании в днешно време осъзнават, че за да оцелеят в съвременната постоянно променяща се среда е необходимо да разполагат с интегрирана и качествена информация за вътрешната и външна околна среда, така че да имат възможност да следят постоянно собственото си положение на пазара. Чрез тази информация, те ще могат да намалят разходите си, ще могат да рационализират различните процеси в компанията и ще подобрят бизнес показателите си.

Традиционните системи вече не успяват да отговорят на нуждите на бизнеса и да предложат приемливо решение за цялостното управление на данните в организацията. Така много от компаниите стигат до извода, че е необходимо да създадат отделен Склад от данни, който да интегрира информацията от различните системи в една единна база данни.

Защо традиционните системи не са достатъчни и защо компаниите трябва да създадат отделен Склад от данни и СУБД? Именно този въпрос ще бъде разгледан теоретично, а след това ще бъдат приведени и доказателства с помощта на нагледен пример.

II Теоретична част:

1. Терминология използвана в настоящата курсова работа:

- Бази данни
- СУБД
- BI
- OLTP
- OLAP
- ETL
- Data Warehouse
- Data mart

2. Проблеми на традиционната файлова организация на данните

Традиционните системи не предлагат приемливо решение за цялостното управление на данните в организацията, поради редица недостатъци. Те са създадени независимо една от друга в различните отдели и от различни хора, поради което данните се интерпретират от позицията на съответния отдел, а не от гледна точка на цялата организация. При традиционните подходи всяка една функционална единица в предприятието създава свое собствено приложение. Всяко едно от тези приложения

изисква свои собствени файлове и отделна компютърна програма, на която да работи. Например, ако разгледаме отдел Човешки ресурси, той ще има няколко файлове: един основен файл с лични данни, файл за здравно осигуряване, за пенсионно, друг файл с майли и други данни и т.н. Така в компанията ще бъдат създадени и поддържани множество основни файлове от всеки отдел. След около 5 или 10 години, компанията ще се окаже с множество системи и приложения, които освен, че ще се поддържат много трудно, най-вероятно ще съдържат различни данни и информация. Резултата от всичко това ще бъде несъответствие и липса на данните, ще липсва сигурност по отношение управлението им и няма да може да се споделя информация между различните приложения и системи.

Друг недостатък на тези системи е че данните ще се дублират и едни и същи данни ще съществуват на повече от едно място. Това съответно ще води до загуба на ресурси и несъответствие в информацията.

Поради наличието на толкова много приложения и системи, много трудно би могло да се осъществи контрол от мениджмънта, който има достъп до съответните приложения и който не.

При описаните традиционни подходи се наблюдават и различия в разбиранята на потребителите за информацията и данните, тъй като те имат различен източник. Това от своя страна води до спорове между служителите, което от своя страна пък води до загуба на време. Като резултат от това ще се създават и различни отчети в организацията, което в днешната динамична среда би могло да е пагобно за една организация.

3. Управлението на данни с помощта на БД подхода.

Технологията на БД, може да се справи с много от проблемите на традиционната файлова организация на данните. Една от основните дефиниции за БД е сбор от данни, организирани да обслужват множество приложения чрез централизиране на данните и контролиране на липсваща или повтаряща се информация. Данните не се съдържат в различни файлове за всяка една апликация, а се съхраняват на едно единствено място. Една единствена база данни обслужва множество приложения. Така например вместо корпорацията да съхранява информацията за служителите в отделни информационни системи и отделни файлове – за лична информация, за заплащане и др. може да бъде създадена една единствена база данни за отдел Човешки ресурси, съдържащ цялата тази информация на едно място.

4. Системи за управление на Бази Данни

Системите за управление на Бази данни представляват софтуер, който позволява на организацията да централизира данните, да ги управлява ефективно и да осигурява достъпа до данните чрез специализирани програми. СУБД играе ролята на нещо като интерфейс между приложенията и физическите файлове с данни. Когато приложението

извика някакъв елемент на данните, СУБД намира този елемент в базата от данни и го предоставя на приложението. При традиционните файлови подходи, това щеше да се осъществи, като се специфицират размера и формата на всеки търсен елемент от данните използван в програмата и след това каже на компютъра къде се намират тези данни.

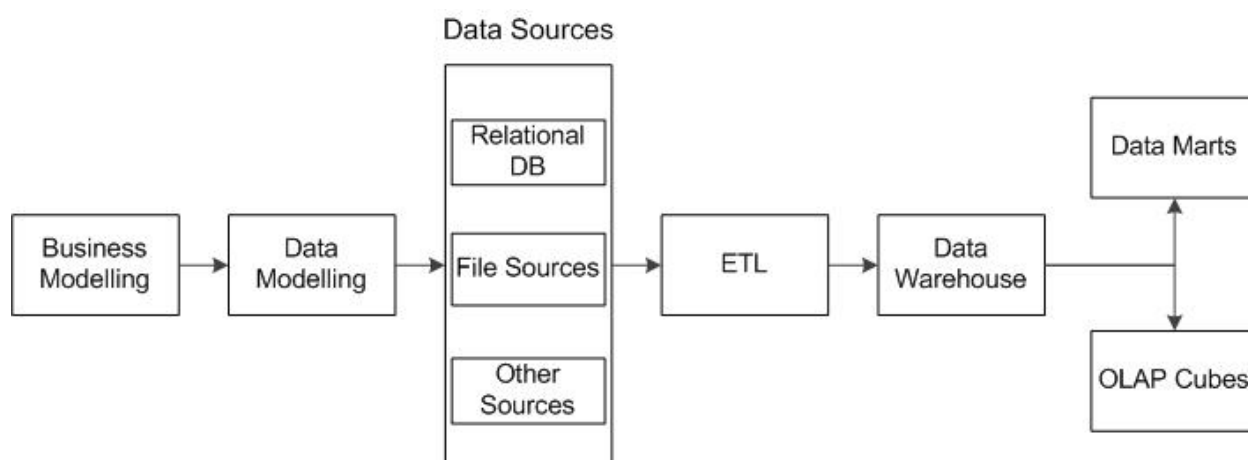
Системите за управление на бази данни освобождават програмиста от задачата да разбира къде и как данните са складирани, като разделят физическия от логическия изглед на данните. Логическия изглед на данните ги представя така че те да могат да бъдат използвани от крайния потребител или бизнес специалиста, докато физическите показват как всъщност са структурирани данните върху физическия носител. Логическото описание на цялата база данни показваща всички елементи на данните и връзките между тях се нарича концептуална схема на данните, докато спецификациите относно как данните от концептуалната схема се съхраняват на физическия носител се представят чрез физическа схема.

5. Компоненти на една Business Intelligence среда.

Целта на Business intelligence е именно превръщането на данните на фирмата в полезна информация. Имплементиране на BI е дълъг процес, който изисква множество анализи и инвестиции. Една типична BI среда включва бизнес модели, модели на данните, източници на данните, ETL, приложения необходими да трансформират и организират даните в използвана и полезна информация, Склад от данни, data marts, OLAP анализи и приложения за създаване на отчети.

Ще разгледаме всеки един от тези елементи и неговата роля в BI средата.

На Фиг.1 по-долу може да се види примерна BI среда.



Фиг.1

5.1 Бизнес моделиране

Бизнес модела представя една обща картина на бизнеса, какво представлява той, какъв бизнес проблем следва да решава той и какъв е потокът на информация от източника до крайната дестинация. Бизнес моделирането включва бизнес

модели и диаграми, които осигуряват информация в графичен вид на членовете на организацията да разбират бизнес правилата и процесите ефективно. Моделирането на бизнес процеси, моделирането на потока от процесите и моделирането на потока на данни са подкатегории на бизнес моделирането. При бизнес моделирането не е необходимо да се навлиза в детайли и много често няма да се разгледат средствата за постигането на задачите. Стратегиите за бизнес моделиране в различните организации се различават в зависимост от техните нужди и цели.

- Моделирането на бизнес процеси представлява група от свързани дейности или бизнес процеси. По-надолу в настоящата курсова работа може да се види именно такъв бизнес модел.
- Моделирането на потока на процесите – графично се описват различните процеси, които се случват в една организация и връзката между тях.
- Моделиране на потока на данни – подход, който описва потока от данни данни между различните бизнес процеси и помага да се обяснят и документират движенията в една бизнес организация

5.2 Моделиране на данните

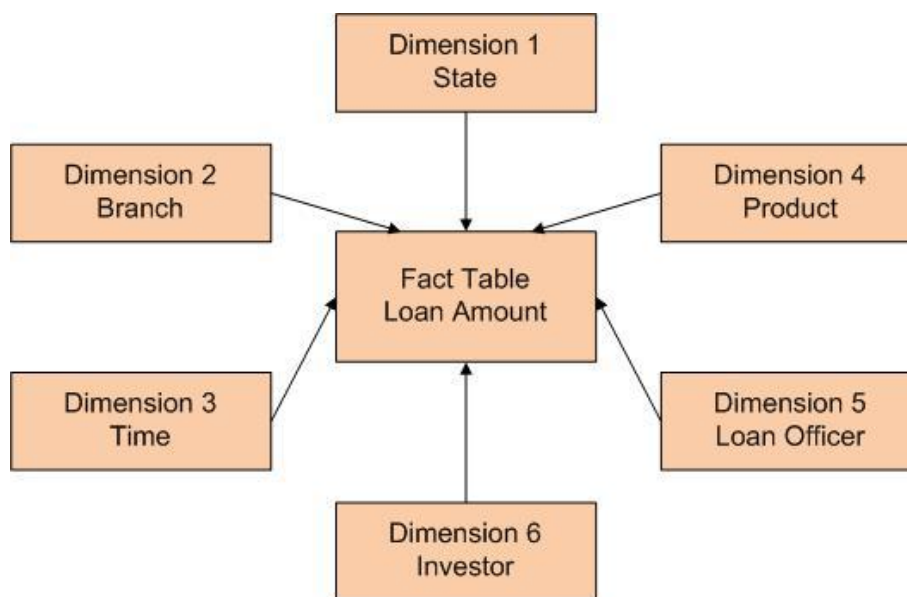
Моделирането на данните представя структурата на данните и техните връзки в моделите от данни, които се включват в една БД. По-просто казано, моделирането на данни представлява визуализация на базата данни. Моделирането на данните се състои от няколко фази като създаването на концептуален модел, създаване на логически модел, създаването на физически модел, създаване на модел от данни на предприятието, релационен модел и дименсионен.

- Концептуален модел: визуализира структурата на базата данни и осигурява обща информация за нея, т.е. не съдържа много детайлна информация.
- Логически модел: представлява разширение на Концептуалния модел и включва почти всички атрибути и техните връзки. Този модел определя бизнес информацията и правилата и не съдържа информация за атрибутите като напр. Тип, дължина и т.н.
- Физически модел: Физическия модел съдържа структура на данните като таблици, колони и връзките между тях.
- Модела на данните на предприятието: познат е като глобален бизнес модел, който консолидира информацията в цялата компания

- Релационния модел: този модел е предпочитан за OLTP процесите. При него таблиците са нормализирани с цел да се избяга повторението на информация.
- Дименсионния модел се състои от една факт таблица и много дименсионни таблици и се използва за пресмятането на обобщени данни. Дименсионния модел е предпочитан в BI средата. Факт таблицата включва например различни мярки като сумата от продажбите, сумата от заемите, докато дименсионните таблици описват конкретна единица като време, регион и др. , на базата на които се правят измерванията на фактическите данни.

В Дименсионния модел, една фактическа таблица е централизирана таблица, която се свързва с различни дименсии. Този подход е наречен “Star Schema”, на основата на който се построи Data Warehouse и Data marts.

На схемата по-долу може да видите примерна такава диаграма от типа “Star Schema”, която описва как може да се пресметне сумата от заемите на база регион, продукт, клон, време и други.



Фиг. 2

5.3 Традиционната “транзакционна” информация генерирана при функционирането на някои от секторите като например продажби на дребно, финансови услуги, телекомуникации, транспорт или казано по друг начин обработката и записването на оперативните данни, които документират ежедневната работа на една компания са известни като “online transaction processing” или OLTP.

OLTP включват обработката и записването на транзакции, които създават нови данни или обновяват съществуващата информация в Оперативните БД: вмъкване (insert), обновяване (update), изтриване (delete). Всяка транзакция обикновено манипулира малък брой записи. Традиционните СУБД са оптимизирани да функционират добре в OLTP обработката, но не и в задачи за обширни анализи, обобщения и създаване на решения. Информационната система, поддържаща ежедневното опериране на бизнеса, базирана на OLTP, се нарича Оперативна система за обработка на транзакциите (Operational Transaction Processing System- OTPS). Дизайна им е съобразен за въвеждане и редактиране на данни в реално време. Примери за такива OTPS системи могат да бъдат: приложения за проследяване на поръчките, приложения за обслужване на клиентите, point-of-sale приложения и др.

OLTP системите служат като източник на данни за Склада от данни, като данните се извличат, трансформират и след това зареждат в Склада от данни посредством ETL процесите. По-долу може да се види подобна графика, която описва тези процеси.

5.4 ETL, Data Warehouse, Data mart

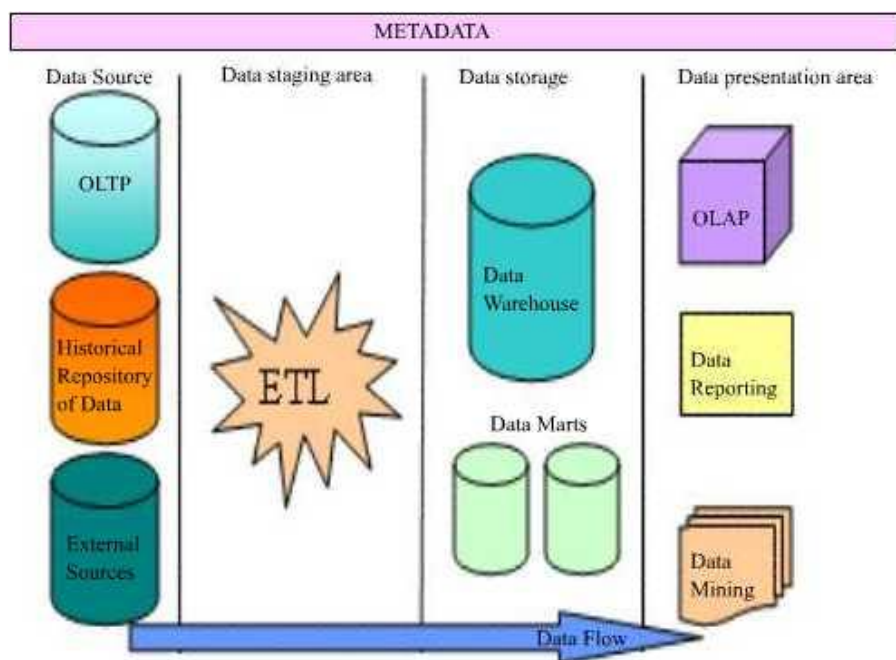
ETL, което е съкращение от “Extract, Transformation, Load” представлява сбор от процеси включващи извличането на данните от източниците, трансформацията на тези данни и най-накрая зареждането на данните в Склада от данни. Преди данните да бъдат заредени в Склада от данни, те трябва да бъдат трансформирани, така че да отговорят на нуждите на Склада от данни. Този процес включва редица други процеси като: почистване на данните (data cleansing), профилиране на данните, конвертиране на типа на данните, денормализация и нормализация.

Склада от Данни съдържа данните събирани от различните източници, с цел анализи и отчети. От Склада от данни могат да бъдат извличани различни отчети с помощта на тулове и data marts.

Data mart-а представлява част от Склад от данни и обслужва конкретна линия от бизнеса или отдел. Той съдържа информацията специфична за определен област, напр. Продажби, доставки или други. Data мартовете могат да бъдат създадени като част от Склада от данни, но могат да съществуват отделно и създадени директно от източника на данни. Те също като Склада от данни са структурирани на базата на дименсионните модели.

5.5 OLAP, или “Online Analytical Processing” е софтуерна технология, която визуализира и обобщава данните, извлечени от Склада от данни и Data mart-а с цел да представи една по-обща картина спрямо различни измерения. Обикновено OLAP се представя чрез различни OLAP тулове (като напр. Cognos, Business Objects и др.), които дават средство на анализаторите, ръководителите, и топ-мениджърите за лесен и ефективен достъп до и анализ на данните. Тъй като Склада от данни е организиран на база дименсионен модел, данните се представя като кубове, които позволяват агрегация на фактите, които да бъдат разделени спрямо различните измерения. OLAP туловете дават възможност данните да могат да бъдат представени на различни йерархични нива.

Средствата за разработка на OLAP анализите са графични и имат възможности за представяне на данните в разнообразни изгледи, отразяващи бизнес информацията и по този начин тя се разбира по-добре от потребителите.



Фиг. 3

III Практическа част:

Управлението на данни в днешно време играе важна роля в областта на финансовите услуги и банките. Време е тези институции да премислят и внедрят новосъздадените решения в областта на Data Warehouse и Data Management.

Финансовата криза от 2008 година показва ясно, че финансовите институции зависят силно от качеството на аналитичните данни и информация. Във връзка с това възникнаха много нови изисквания и подходи спрямо центровете на данни на фирмите и възможностите за качествени анализи на показателите. Традиционните подходи спрямо Data Warehouse и Анализите, които бяха актуални по време на кризата, не успяха да отговорят на очакванията в един критичен период. Много от тези традиционни подходи обаче все още продължават да се прилагат от компаниите.

Така, в същото време обаче, състезанието между банките, увеличава напрежението им да подкрепят своите операции с аналитична информация, която от своя страна да бъде доставяна на време за оперативното вземане на решения. Това е в силен контраст с традиционните системи, където е характерно „offline” вземане на решения, което е в полза единствено на висшия мениджмънт.

В днешната вече трансформирана среда, много фирми започват да се събуждат и да преосмислят нуждата от нови стратегии в областта на Управлението на данни.

Ето защо поради актуалността на темата, за нашия казус ще изберем примерна финансова институция “X” и ще разгледаме един от процесите изпълнявани в нея. Ще разгледаме как по-точно внедряването на Data Warehouse и OLAP системите помага за правилното и бързо решение на проблемите.

Ще опишем процес във финансова организация, малка част от всички изпълнявани в една организация процеси, с цел да се обясни по-нагледно какъв е потокът на данни в една фирма, как се събират данните, в последствие как те се трансформират и обработват с цел да могат да се използват за анализи и вземане на решения.

Ще разгледаме финансова институция “X”, която се занимава с потребителско кредитиране. Един от първоначалните процеси извършвани във фирмата е процесът по Кандидатстване за кредит. Стъпките описани в Бизнес модела по-долу са задължителни за изпълнение при кандидатстването за нов кредит.

Първоначално Специалистите Първо ниво получават заявление за получаване на кредит от офисите или магазините, в които се съдържа необходимата информация на клиента. Те разглеждат тази информация, като осъществяват проучване на клиента и решават дали този клиент може да бъде одобрен или не, в зависимост от определени изисквания на които той трябва да отговаря. Ако заявката бъде одобрена на първо равнище, се разглежда и от Специалист Второ ниво, който отново проверява дали клиента отговаря на изискванията. Проверява се състоянието на клиента в Централен Кредитен Регистър (ЦКР) на БНБ, с цел да се види дали има някакви неизплатени кредити към други финансови институции и с какъв статус са тези кредити. Специалист Второ ниво проверява и резултата за разглеждания клиент от предварително вътрешна създадена скор карта на база подадените данни. След осъществяването на всички тези проверки се решава дали на клиента ще бъде отпуснат кредит или не.

В последствие се осъществява контрол по отпуснатите кредити, правят се различни анализи, отчети и справки на база на които се вземат управленски решения по отношение на продажбите и предлаганите продукти. Създават се скор карти от Риск отдела, както и различни анализи, които проследяват колко редовно се плащат вноските по кредитите, колко от тях закъсняват и с колко време.

Въпросът е как трябва да бъдат организирани и структурирани данните на фирмата, за да могат да бъдат осъществени всички тези задачи. Дали традиционната система би могла да отговори на нуждите на бизнеса? С оглед на информацията, която ни е необходима за проследяването на тези показатели и индикатори, а в последствие и построяването на различни трендове, които да представят движението им във времето е необходимо създаването на Склад от данни. Склада от данни ще извлича информацията от OLTP системите посредством ETL процеси, и в последствие посредством OLAP тулове ще визуализира и обобщава данните. В случая, ако приемем, че разглежданата от нас финансова институция “X” не е толкова голяма, може да си спестим разходите за създаването на Data mart и OLAP туловете да извличат информацията директно от Склада от данни.

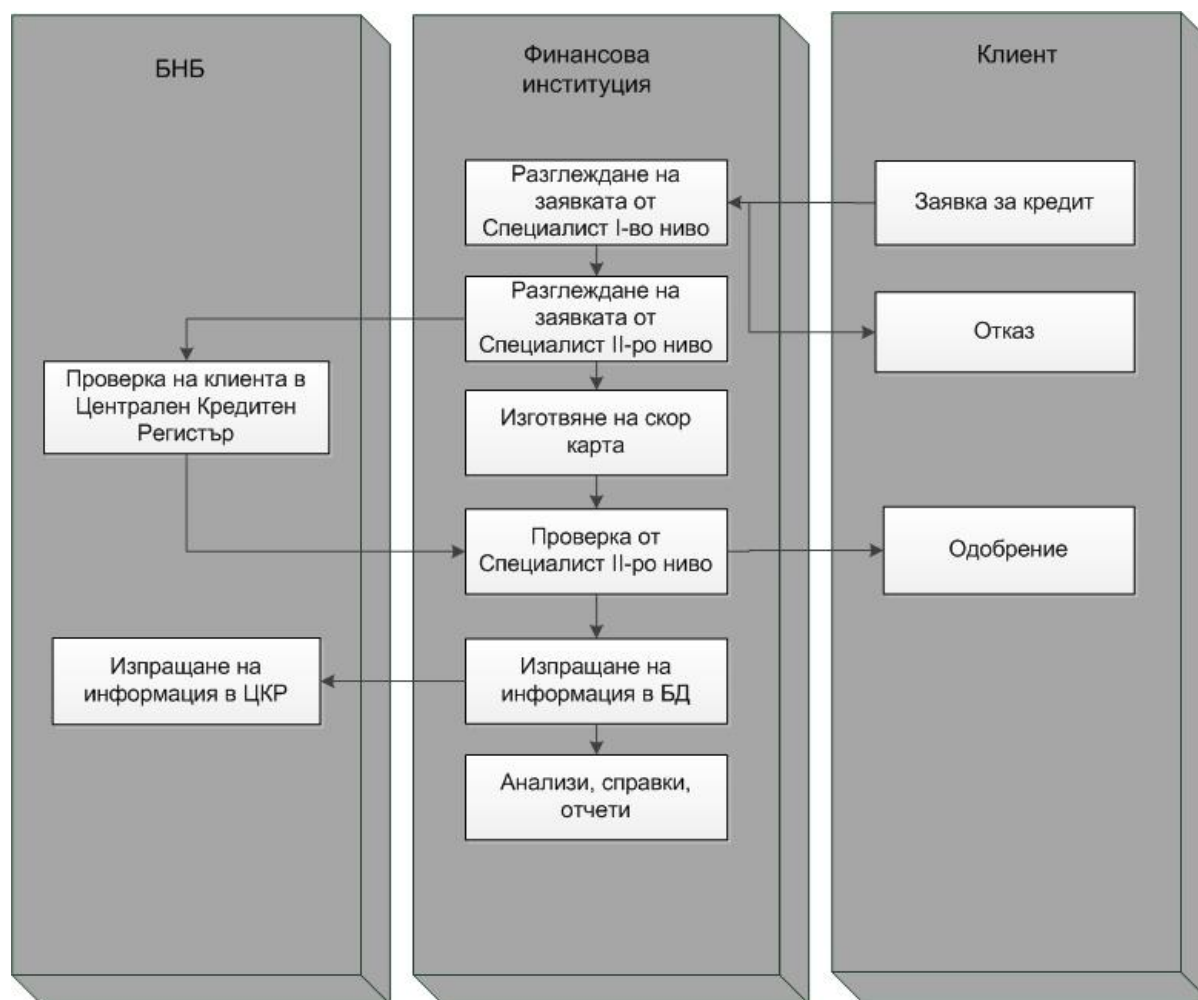
Необходимо е също така да се проследява текущия статус по кредитите и състоянието да се изпраща в Централен Кредитен Регистър за обща информация. Описаното е една малка част от нуждите на съответната компания. Необходима е много друга допълнителна информация, която да послужи пред Управленския състав, за да може да съществува нашата финансова институция и да е на печалба.

Така решението на представения казус в настоящата курсова работа ще бъде в четири части: Business view, Process View, Applications View и Technical View, където е направен опит да се представи схематично решение на поставения казус.

Ще представим схематично и алтернативен вариант на поставения проблем, т.е. решение на казуса без наличието на Склад от данни.

1. Business view

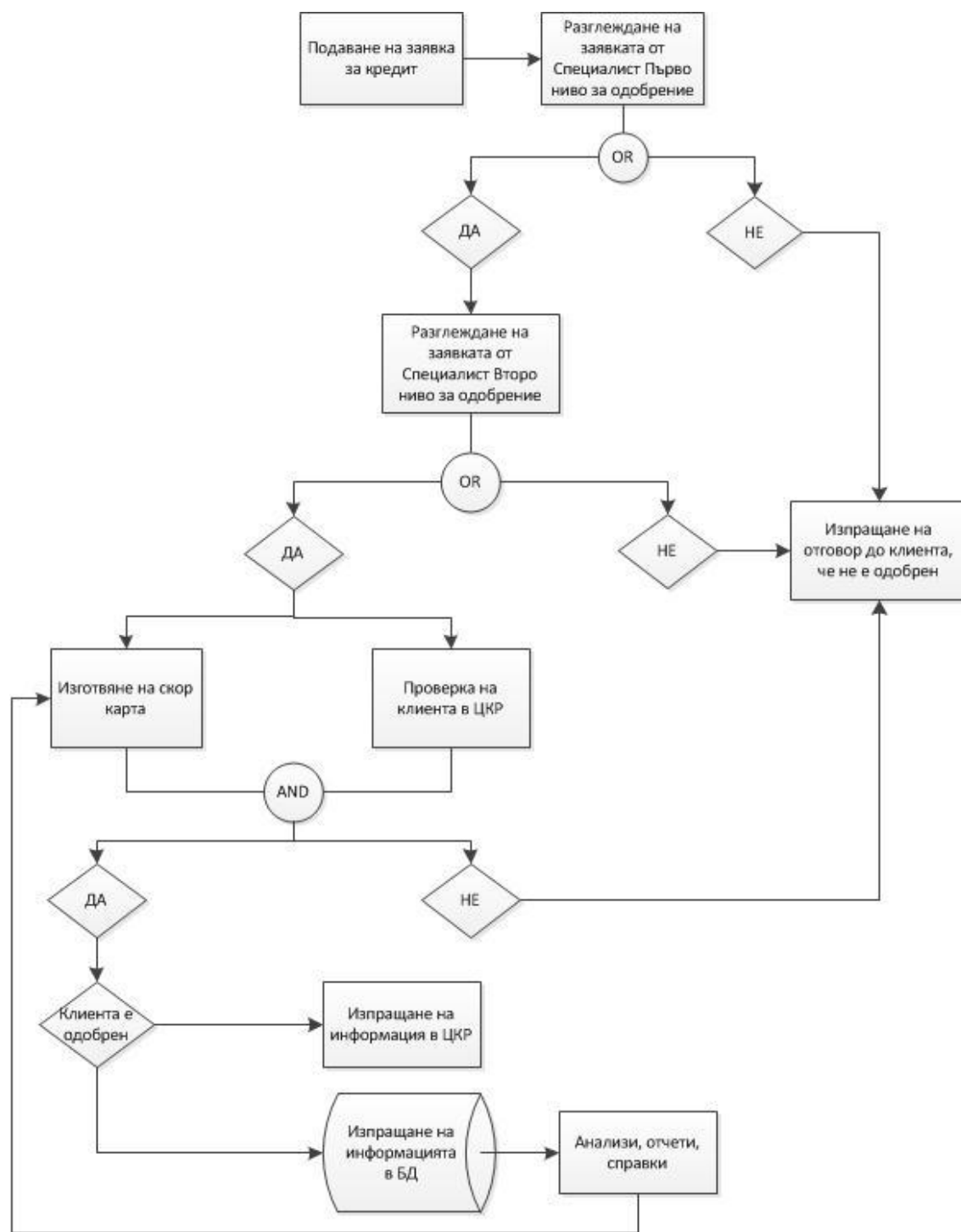
Тук именно е представен бизнес модела, който представлява група от свързани дейности или бизнес процеси. В нашия случай са представени връзките Клиент-Финансова институция-БНБ и процесите, които протичат между тях.



Фиг. 4

2. Process view

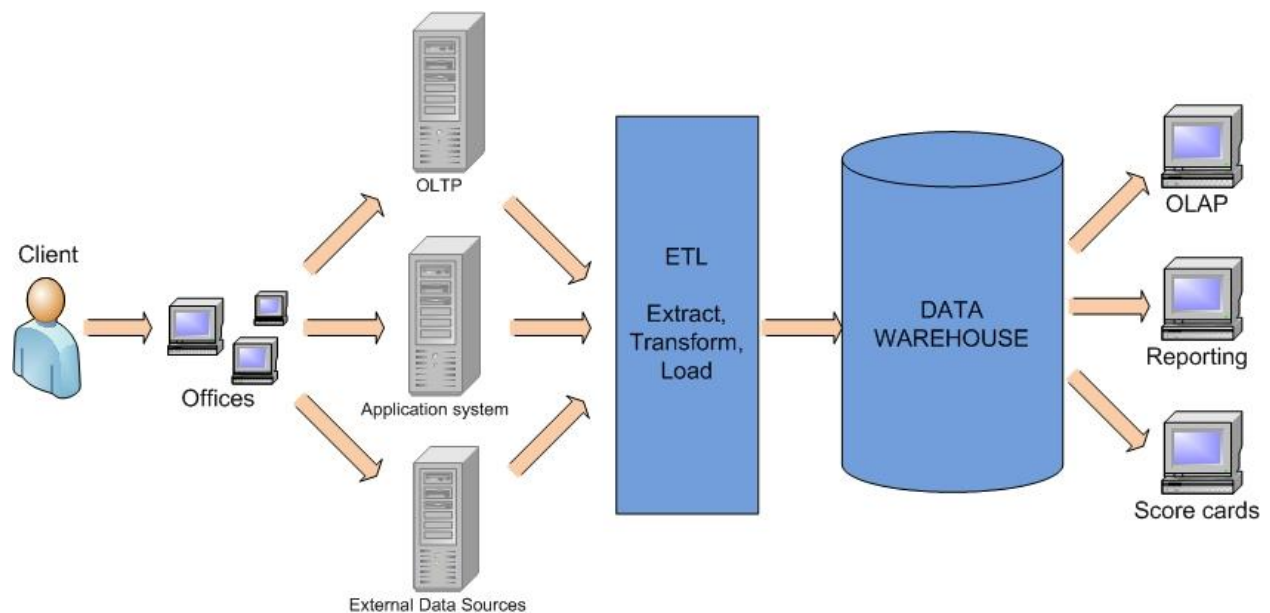
Процесния модел, който описва графично различните процеси, които се случват в една организация и връзката между тях. Представен е един от многото такива процеси, които протичат в нашата финансова институция.



Фиг. 5

3. Applications view

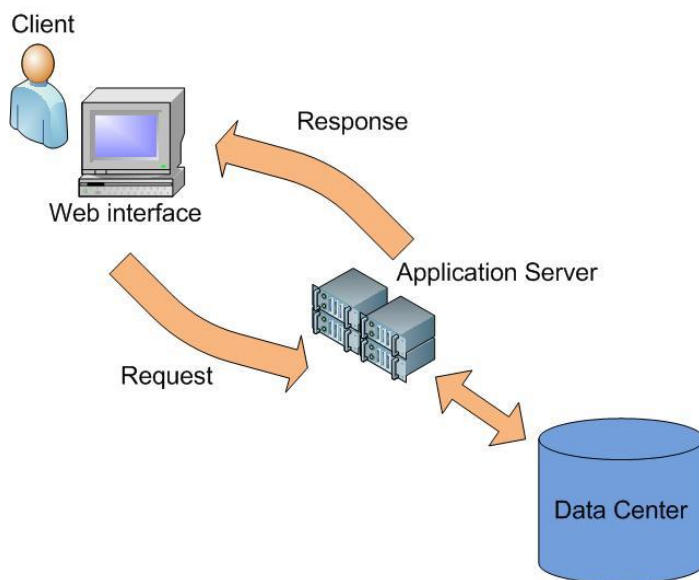
Тук може да бъде видян потока от данни между различните бизнес процеси и приложения в нашата компания. Това помага да се обяснят и документират движенията в една бизнес организация.



Фиг. 6

4. Technical View

На Фиг.7 може да се види техническият изглед на разглеждания казус. През web interface се подават исканията за кредит, разглеждат се и се одобряват. На Application ниво се правят скоринг анализите и различните необходими проверки, за да бъде взето решението дали кандидатът е одобрен или не.

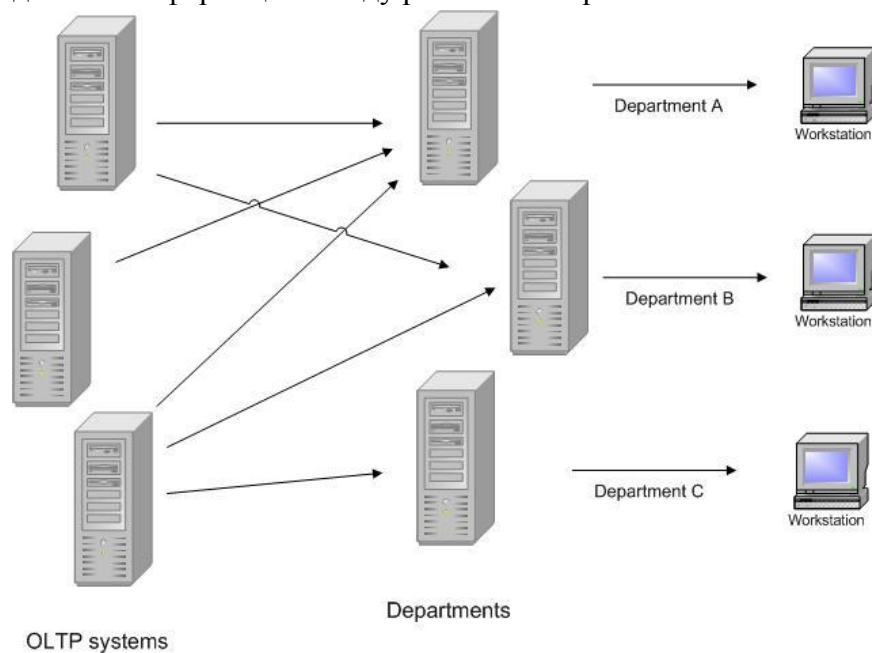


Фиг. 7

5. Алтернативен вариант

Как би изглеждало описанието на бизнес процесите, без наличието на Склад от данни и какъв ще бъде потокът на информация в този случай може да се види на схемата по-долу.

Тук добре се вижда, че по този начин ще съществува проблема с липсата на централизирана информация, данните ще съществуват на повече от едно място, което със сигурност ще доведе до дублирането им. Също така няма да има възможност за споделянето на данни и информация между различните приложения и системи.



Фиг. 8

Управлението на данни както разбрахме е от изключително значение за бизнеса в съвременната постоянно променяща се среда, и то не само в областта на финансовите услуги, но и в много други области като например здравеопазване, телекомуникации, инвестиционни и осигурителни компании и др.

Нуждата от Склад от данни е налице и най-вече заради следните условия:

- Всяка компания трябва да познава индустрията и пазара на който оперира.
- Трябва да може да посреща предизвикателствата, като например промени на изискванията към продуктите или услугите.
- Трябва да инвестира в изследователски програми за нови продуктови линии и нови технологии.
- Да инвестира в доходоносните клиенти, които носят печалби.
- Да се опита да запази настоящите си клиенти и да привлече нови.
- Да се стреми да повишава достъпа до информация, нужна за вземане на бързи и правилни решения.

За всичко това е необходимо превръщането на данните в надеждна и полезна бизнес информация, с цел анализи и вземане на правилни и бързи решения.

Източници на информация:

1. Management information systems – Management the digital firm
2. Inmon - Building The Data Warehouse 3rd Ed [Wiley 2003]
3. www.banktech.com/business-intelligence/232400413
4. www.learnbi.com/bi_mining.htm